Adaptive Paywall Mechanism for Digital News Media

Heidar Davoudi
York University, Canada
davoudi@cse.yorku.ca

Aijun An
York University, Canada
aan@cse.yorku.ca

Morteza Zihayat
Ryerson University, Canada
mzihayat@ryerson.ca

Gordon Edall
The Globe and Mail, Canada
GEdall@globeandmail.com

ABSTRACT

Many online news agencies utilize the paywall mechanism to increase reader subscriptions. This method offers a non-subscribed reader a fixed number of free articles in a period of time (e.g., a month), and then directs the user to the subscription page for further reading. We argue that there is no direct relationship between the number of paywalls presented to readers and the number of subscriptions, and that this artificial barrier, if not used well, may disengage potential subscribers and thus may not well serve its purpose of increasing revenue. Moreover, the current paywall mechanism neither considers the user browsing history nor the potential articles which the user may visit in the future. Thus, it treats all readers equally and does not consider the potential of a reader in becoming a subscriber. In this paper, we propose an adaptive paywall mechanism to balance the benefit of showing an article against that of displaying the paywall (i.e., terminating the session). We first define the notion of cost and utility that are used to define an objective function for optimal paywall decision making. Then, we model the problem as a stochastic sequential decision process. Finally, we propose an efficient policy function for paywall decision making. The experimental results on a real dataset from a major newspaper in Canada show that the proposed model outperforms the traditional paywall mechanism as well as the other baselines.

CCS CONCEPTS

• **Theory of computation** → *Sequential decision making*;

KEYWORDS

Digital news media; Paywall; Sequential decision making

ACM Reference Format:
Heidar Davoudi, Aijun An, Morteza Zihayat, and Gordon Edall. 2018. Adaptive Paywall Mechanism for Digital News Media. In *KDD '18: The 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, August 19–23, 2018, London, United Kingdom.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3219819.3219892

ACM ISBN 978-1-4503-5552-0/18/08…$15.00
https://doi.org/10.1145/3219819.3219892

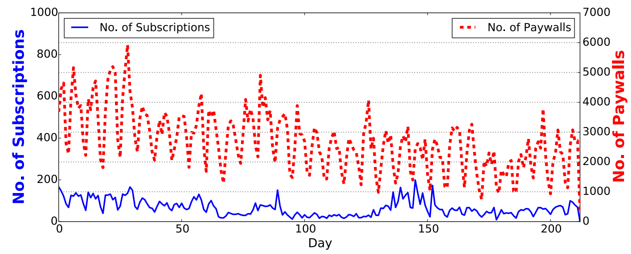


Figure 1: The number of subscriptions vs. the number of paywalls in The Globe and Mail dataset for the period 2014-01 to 2014-07. There is a weak correlation (i.e., $\rho = 0.59$) between the two numbers.

1 INTRODUCTION

Most online newspapers across the world generate revenue by displaying advertisements or/and using a pay model that restricts the reader access to articles via a paid subscription. In the former one, news agencies (e.g., USA Today) operate based on an ad-supported free content model, in which the articles are accessible for free and the revenue is derived from displaying advertisements. However, advertisement revenues may not be sufficient to sustain existing forms of news production as they do not create long term relationships with customers [2, 9]. Therefore, pay models, also known as *paywall mechanisms*, were developed by digital media to increase revenues through subscription. In such models, news agencies (e.g., The New York Times, and The Globe and Mail) offer a certain number of free articles in a period of time (e.g., a month) and then redirect visitors to the subscription page (i.e., paywall) to continue reading articles. The ultimate goal of a paywall mechanism is to persuade users to subscribe and as a result boost the profit. However, user persuasion for subscription (i.e., user acquisition) is not an easy task in news domain since users usually have many choices in selecting news sources. Moreover, in most cases there is no direct relationship between the number of paywalls presented to readers and the number of subscriptions. That is, by increasing the number of times that paywalls presented to readers, we may not necessarily raise the number of subscriptions (see Figure 1). Therefore, the traditional paywall mechanism based on the total number of articles read in a period may not serve its purpose of increasing revenue. It may actually turn away many potential subscribers.

Due to the increase of technological obstacles for ad-supported free content revenue models (e.g., ad blockers) and the number of online news providers seeking for sustainable relationship with

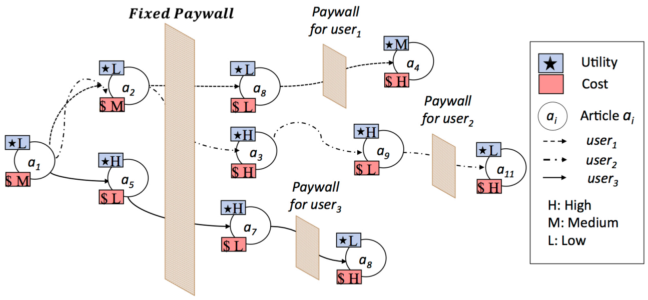


Figure 2: Incorporating utility and cost in paywall mechanisms.

customers (78 % of U.S. newspapers with circulations over 50,000 are using a digital subscription model [18]), the need to generate revenue by developing effective paywall mechanisms is demanding. This is due to the fact that blocking a user from reading articles at a wrong time may disengage the user too early or allow a non-potential subscriber to read too much content for free. Therefore, developing a smart paywall policy is of paramount importance to the prosperity and profitability of an online newspaper. The availability of users' interaction data and advances in machine learning techniques raise an interesting question: *Can we estimate how many and what articles a particular user should be allowed to read before a paywall?* It is unrealistic to expect the same answer for all users. Moreover, the answer should consider business objectives even if they could be in conflict with one another. For example, allowing readers to read more articles leads to more display of ads, increasing the ad-based revenue. However, from the subscription point of view, this is not desirable as offering too much free content makes subscription unnecessary from the reader's point of view.

Finding the optimal paywall time is a sequential decision-making process (or a sequential decision problem), in which a decision whether to show a paywall needs to be made at each time point during a reading session and once the paywall is presented, the session terminates. To define an objective function for this problem, we introduce the notion of *utility* and *cost*. The utility of an article measures the effectiveness/usefulness of the article in achieving a business objective (e.g., user engagement which can increase the subscription possibility). The cost of an article measures the amount of resources consumed to prepare it (e.g., the amount of money paid to the author). Traditional paywall models (called *fixed paywall* or *metered paywall*) block users from reading articles after visiting a certain number of articles (e.g., two articles) without considering the above two or any other factors. Figure 2 shows a toy example demonstrating how the utility and cost of an article can be used to make smarter decisions. Assume that $user_1$ and $user_2$ both visit two low-utility articles a_1 and a_2 (e.g., ones that can be found in many other news sources). If the fixed paywall policy with the limit of two articles is used, both of them would be directed to the paywall. Assume we do not use the fixed paywall but consider what they would like to read next. Suppose $user_1$ clicks on a_8. Since a_8 has a low cost, we can show it to the user and take advantage of other benefits (e.g., displaying ads when showing a_8). However, $user_1$ might not be a good target for subscription as all the articles she has read have a low utility (e.g., can be found somewhere else), so once she clicks on the next article (i.e., a_4), which has a high cost, the paywall is presented. On the other hand, $user_2$ can see the article a_3 as its high cost can be justified by its high utility, and then receives the paywall after visiting a_9, where the next article the user selects has low utility and high cost (i.e., a_{11}). Moreover, a user (e.g., $user_3$ in Figure 2) who visits high utility articles (i.e., a_5, a_7), which may be articles visited by subscribers before their subscription, is more likely to become a subscriber. Therefore, we show the paywall (after a_7) since the next article (e.g., a_8) has a high cost and low utility. Note that this user is more likely to become a subscriber and presenting the paywall at the right time might persuade her to subscribe. This example shows that different reading behaviors may require different paywall strategies, and making decisions based on the utility and cost model can serve as an effective approach to make a smart decision per user visit.

However, finding the optimal paywall time for a user is a challenging problem. First, while the concepts of utility and cost provide insights into the paywall decision problem, incorporating them into the optimal decision making process in a disciplined way is not a trivial task. Second, when making a decision at a time point (i.e., when the user clicks on an article), we only know the articles that the user has visited and the next article she is trying to read. The articles beyond the next one are unknown. Thus, when making an optimal decision at an early point, we need to consider the uncertainty as to what would happen later. Third, a proposed model is better to be flexible so that any objective can be plugged in as desired. Last but not least, the proposed approach should be efficient to work in an online setting.

To address these challenges, we formulate the problem in a unified stochastic decision making framework by considering the utility and cost of articles. We define the main components of the model, propose an effective approach to solving the problem, and provide theoretical supports for the proposed approach accordingly. The main idea is that at each stage the paywall will be presented to a user if the prospective articles (which are likely to be visited by the user) are not promising in terms of the utility and cost.

Our main contributions are as follows:

- We define the new problem of adaptive paywall mechanism in digital news media. While this problem is a major issue in subscription-based online news agencies, to best of our knowledge, it has not been studied before.
- We cast the problem into a sequential stochastic optimization problem in a disciplined way, and propose an effective data-driven solution accordingly. In particular, we provide the theoretical analysis and design an effective policy for the problem. The proposed framework is general in that it can be applied with any given business objective.
- We apply the proposed framework to a real dataset obtained from a major Canadian newspaper and show that it outperforms some baseline approaches in terms of different business objectives.

The rest of the paper is organized as follows. Section 2 discusses the related work. The problem and framework components are defined in Section 3. We describe the proposed model in Section 4.

Section 5 presents an application of the proposed method and its empirical evaluation, and finally Section 6 concludes the paper.

2 RELATED WORK

The founding of journalism has been a major issue for the news industry over the past decade [10]. While diminishing income due to decline in the advertisement revenue makes newspapers to start implementing paywall mechanisms, there are few studies on the analytical side to make this mechanism more effective. Most studies in the journalism community focus on the qualitative and quantitative investigation of features (e.g., age) influencing people to pay for subscription [4, 5, 7, 8].

The sequential decision making over time has been studied in many disciplines with different names such as: *reinforcement learning* in machine learning, and *approximate dynamic programming* in operation research. In reinforcement learning, Markov Decision Process (MDP) is widely used to model the dynamics of an environment under different actions (i.e., decisions). When the environment model is available, dynamic programming methods such as *value iteration* or *policy iteration* can be used to find the optimal policy [14]. For example, Cai et. al [3] considered the biding problem where the main goal of the advertiser is to bid for every ad impression in an auction. Given a bid request, in each timestamp the advertiser should make a decision based on the ad request contexts and its current state (e.g., amount of budget). They designed a method based on Markov Decision Process (MDP) to learn the optimal decision policy. However, the environment model is often not available. In such cases, Temporal Difference learning [13] techniques such as Q-learning [17] or SARSA [14] can be utilized to learn the policy from the environment. However, model free approaches need a lot of interactions (i.e., exploration) with the environment before the convergence and suffer from transition dynamics of an enormous state space and the sparsity of reward signals in the highly stochastic environment. Shani et. al [12] considered the recommendation (in the bookstore domain) as a decision problem compared to the traditional prediction perspective. They designed a framework based on MDP and utilized the value iteration to make an optimal decision (i.e., whether to recommend an item to a user or not). However, due to the limitation on exploration they finally used some heuristic techniques to learn the policy.

Approximate dynamic programming studies the sequential decision problem in a more general setting and with broader policy classes such as *myopic policy* (i.e., decision is based on the current state) or *lookahead policies* (i.e., decision is based on the predicted decision in the future) [11]. One related problem in this area is American price optioning [1], which has been studied extensively. American option allows option holders to exercise the option before the maturity date. This problem has been studied in the the stock dynamics in the risk neutral world (i.e., a stochastic differential equation). Despite some similarities between this problem and the paywall problem, in the paywall problem we do not have such a dynamic instead we need to build the solution in a data-driven fashion.

3 PROBLEM DEFINITION

We describe the main components of the proposed model and define the problem accordingly. Figure 3 shows the proposed framework

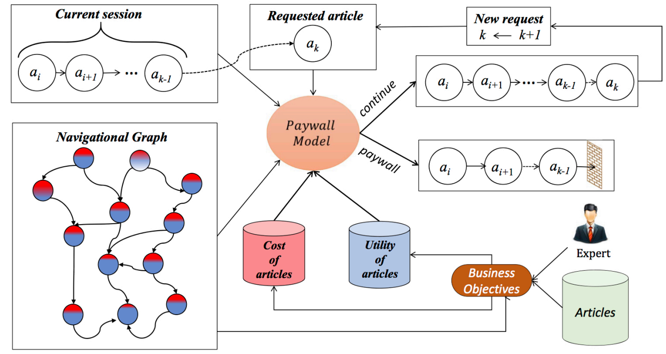


Figure 3: The proposed adaptive paywall framework.

for adaptive paywall. The main components of the framework are: *utility*, *cost* and the *navigational graph* as well as the *paywall model*. The paywall model receives an article request, makes the decision, and changes the current state of the user session accordingly.

Definition 3.1. (**Utility of article**): The utility of article a_i, denoted as $\phi(a_i)$, measures the effectiveness/usefulness of the article in achieving a business objective (e.g., user *engagement* which can increase the subscription possibility).

The utility of an article can be determined by domain experts or learned from historical navigational patterns in a data-driven fashion. For example, if a high percentage of the non-subscribed users reading an article subscribed to the newspaper later, the article has a high utility.

Definition 3.2. (**Cost of article**): The cost of article a_i, denoted as $\psi(a_i)$, specifies the amount of resources (e.g., time, monetary cost) allocated to produce it.

The cost can be specified in different ways. For example, the amount of money which the newspaper has to pay to the author, number of pages, etc.

Definition 3.3. (**Navigational graph**): Navigational graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ is a directed graph, where $\mathcal{V}$ is a set of vertices representing articles, $\mathcal{E}$ is a set of directed edges where an edge e_{ij} from article a_i to a_j indicates a_j has been viewed right after a_i by at least one user, and weight $w_{ij} \in \mathcal{W}$ on e_{ij} represents the number/percentage of the users that read a_j right after reading a_i.

The navigational graph encodes historical user navigation behaviours and is used in our model to estimate what article(s) a user is likely to visit next. Although the graph can be built based on all reading sessions that have occurred, we build the navigational graph based on the sessions made by subscribed users since they do not receive paywalls, and thus less bias is introduced.

A *session* is a group of activities (e.g., requesting and reading an article) that a user spends during one visit to the online newspaper. Traditional paywall mechanisms may consider the information in one or more sessions of an unsubscribed user in making a paywall decision (e.g., presenting the paywall if the total number of articles read by a user exceed a limit for a month). Since an unsubscribed user does not have an ID, user identification for each session (e.g., based on IP addresses or cookies) is necessary in order to consider multiple sessions of a user. However, since different users may use the same IP address and cookies can be blocked, tracking users across multiple sessions may be problematic. Thus, we focus on session-based paywall, although the proposed model can be applied to multiple sessions of a user.

Definition 3.4. (**Session-based Paywall**): In this model, the paywall decision (i.e., whether to present the paywall at a time point in a session) is made based on the information that the user provides in the session without considering historical records of this particular user beyond the session. Once the paywall is presented, the session terminates.

An example of session-based paywall is the traditional fixed/metered paywall that allows a user to read a fixed maximum number of articles (e.g., 2 articles) in a session. In this work, we propose an adaptive session-based paywall model in which the number of articles that a user can read depends on what the user has read/requested in the session and estimations of what the user might read in the future. Our adaptive paywall problem is defined as followed.

Problem Statement (**Adaptive paywall**): Given a navigational graph of subscribed users and a session that an unsubscribed user started, the goal is to determine at what time point during the session the paywall is presented so that the following objective function is maximized:

$$\frac{\sum_{i=0}^{k} \phi(a_i)}{\sum_{i=0}^{k} \psi(a_i)}, \tag{1}$$

where $\phi(a_i)$ and $\psi(a_i)$ are the utility and cost of the ith article that the user reads and $k+1$ is the total number of articles the user reads in the session before the paywall is presented. Although we define the objective function using the utility-to-cost ratio, it can be defined in different ways based on business objectives.

4 PROPOSED METHOD

The adaptive paywall problem is a sequential decision problem, that is, at each time step when the user requests an article in a session, a decision needs to be made regarding whether to allow the user to read the requested article or to present the paywall. Formulating and solving such a problem in a disciplined way is not a trivial task. The major challenge is the huge search space due to the high dimensionality of the problem. At each time step when the user requests a new article, in order to find an optimal solution, we need to look at not only what the user has read and requested, but also what the user would request or read in the future if the paywall is not presented at the current time, in order to compare with the values of the objective function at future time steps. However, what the user will request or read is uncertain at the current time. Considering all possibilities at each time step to find exact solutions is prohibitable due to the large number of articles and combinations of them over multiple time steps. Thus, we resort to approximate solutions. Below we formulate the adaptive paywall problem using the *approximate dynamic programming* paradigm [11] and design a data-driven lookahead policy that makes decisions based on predicted behaviour of the user in the future to solve the problem.

4.1 Proposed Paywall Model

One of the most important tasks in approximate dynamic programming (and in particular, in sequential decision making) is to design a model of the problem, that is, to design the components of the problem. However, despite the importance of this step, there is no standard approach to modeling the problem [11]. A good model can facilitate the design of policy for solving the problem and may also allow the change of the assumptions (e.g., how to combine utility and cost) based on business objectives and providing alternative solutions accordingly. The major components of a sequential stochastic decision problem are: state variable, decision variable/function, transition function, and contribution function. Given the main components of the model the objective function can be defined accordingly.

State variable S_t: The state variable at time step t of a session is defined as $S_t = (\overline{u}_t, \overline{c}_t, a_t)$, where a_t is the *article requested* at time t by the user, and $\overline{u}_t$ and $\overline{c}_t$ are the sum of utilities and costs of visited and requested articles (including a_t) by time t in the session, respectively.

The state S_t encapsulates the accumulated information in a *session* by time t, which is used to make the decision at time t. By defining the state in this way, the next state will not depend on the states before this state, which makes the decision process satisfy the Markov property. Meanwhile, the state representation should contain *minimally* necessary information, otherwise it may make the problem computationally intractable. For example, listing all the articles the user has visited in the session so far in a state results in a much richer representation, but it causes the state space to grow exponentially and makes the problem computationally intractable.

Decision function $X^{\pi}(S_t)$: The decision function determines the decision/action given state S_t using a policy π, where π is a function that maps S_t into a decision/action. In our problem, there are two possible decisions: presenting the paywall or not presenting it. The decision function is defined as follows:

$$x_t \triangleq X^{\pi}(S_t) \triangleq \begin{cases} 1 & \text{if } \pi(S_t) \text{ is ``presenting paywall''} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where x_t is called a *decision variable* indicating the decision taken at time step t. In our framework, policy π is to be designed using a data-driven method.

Transition function S^M: This function depicts the way that the proposed model evolves from one state to another one as a result of decision and exogenous information (i.e., a requested article at the next time step). The transition function S^M determining the transition from state S_t to S_{t+1} given decision x_t is defined as follows:

$$S_{t+1} = S^M(S_t, x_t, \hat{a}_{t+1}) \tag{3}$$

where S^M maps the components of S_t to S_{t+1} as follows:

if $x_t = 0$

$$a_{t+1} = \hat{a}_{t+1} \tag{4}$$

$$\overline{u}_{t+1} = \overline{u}_t + \phi(\hat{a}_{t+1}) \tag{5}$$

$$\overline{c}_{t+1} = \overline{c}_t + \psi(\hat{a}_{t+1}) \tag{6}$$

else

$$a_{t+1} = \text{``}paywall\text{''}$$

where $\hat{a}_{t+1}$ is the article to be requested at time $t+1$. Note that at time t, $\hat{a}_{t+1}$ is uncertain, and thus it is information that arrives exogenously, representing a source of randomness. As a result, its utility $\phi(\hat{a}_{t+1})$ and cost $\psi(\hat{a}_{t+1})$ are random. Also note that if $x_t = 1$ (i.e., the decision at time t is to present the paywall), we assign *paywall* to a_{t+1} to indicate the end state of the decision process. We denote the end/paywall state as S_p.

Contribution function: The immediate contribution/reward function of decision x_t in state S_t measures how much decision x_t at state S_t contributes towards the final objective of the decision process, and is defined as follows:

$$C(S_t, x_t) \triangleq \begin{cases} (\overline{u}_t - \phi(a_t))/(\overline{c}_t - \psi(a_t) + 0.05) & \text{if } x_t = 1 \\ 0 & \text{if } x_t = 0 \text{ or } S_t = S_p \end{cases} \quad (7)$$

Note that when $x_t = 1$ (i.e., when the decision is to present the paywall), a_t is not presented to the user, thus a_t's utility and cost are not included in the accumulated utility and cost when the ratio is computed in the contribution function. 0.05 in the denominator is to avoid zero division. Also, we set the contribution to zero when $x_t = 0$, so that the contribution is only collected at the paywall time because the contribution at the paywall time considers the utilities and costs of all the articles that user reads in the whole session.

Paywall decision problem: We define the paywall decision problem as finding a policy π that maximizes the following objective function:

$$\mathbf{E}\{\sum_{t=0}^{\infty} \gamma^t \, C(S_t, X^\pi(S_t))\} \quad (8)$$

where $\gamma \leq 1$ is a discount factor (emphasizing that contributions in the future is not important as the current time contribution).

4.2 Policy Design

Solving the optimization problem defined in (8) directly is computationally intractable [11]. In this section, we convert Equation (8) into state value functions and analyze different possibilities for the policy design, and in the next section discuss the proposed method.

Let $V^\pi(S_t)$ (called the value of state S_t with respect to policy π) be the expected total contribution of a session starting from state S_t and following policy π. That is,

$$\begin{aligned} V^\pi(S_t) &= \mathbf{E}\{\sum_{t'=t}^{\infty} \gamma^{t'-t} \, C(S_{t'}, X^\pi(S_{t'}))\} \\ &= C(S_t, X^\pi(S_t)) + \gamma E[V^\pi(S_{t+1})|S_t]. \end{aligned} \quad (9)$$

Clearly, Equation (9) is the same as the objective function (8) when using S_t to denote the initial state. Thus, maximizing Equation (8) is equivalent to maximizing (9).

Theorem 4.1. *The optimal value of Equation* (9) *is given by:*

$$V^{\pi^*}(S_t) = \max_{x_t \in \{0,1\}} \{x_t C(S_t, x_t) + (1 - x_t)\, \gamma \mathbf{E}[V^{\pi^*}(S_{t+1})|S_t]\} \quad (10)$$

where π^* is the optimal policy and x_t is the optimal decision at state S_t based on π^* (i.e., $x_t = X^{\pi^*}(S_t)$).

Proof. According to Equation (9) and Bellman's Principle of Optimality [14] that states an optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision, we have:

$$V^{\pi^*}(S_t) = \max_{x_t \in \{0,1\}} \{C(S_t, x_t) + \gamma \mathbf{E}[V^{\pi^*}(S_{t+1})|S_t]\} \quad (11)$$

where $x_t = X^{\pi^*}(S_t)$. We need to show Equation (11) is equivalent to (10). There are two possible decisions in state S_t:

a) If $x_t = 1$, the paywall is presented and thus $S_{t+1} = S_p$ (the end state). Since no contribution can be obtained at the end state and the process ends, $\mathbf{E}[V^{\pi^*}(S_{t+1})|S_t] = 0$. Thus, according to (11), $V^{\pi^*}(S_t) = C(S_t, x_t) = x_t C(S_t, x_t)$.

b) If $x_t = 0$, $C(S_t, x_t) = 0$ according to (7). Thus, according to (11), $V^{\pi^*}(S_t) = \gamma \mathbf{E}[V(S_{t+1})|S_t] = (1 - x_t)\gamma \mathbf{E}[V(S_{t+1})|S_t]$.

The final optimal value based on (11) is the maximum value between case (a) and (b) and obviously can be written by Equation (10). □

Equation (10) provides the insight on how to make an optimal decision for the paywall decision problem. At each state we can make the decision by comparing the optimum value of the state with the expected optimum value of the next states. Note that the value of a next state is computed recursively. One common approach to solve the Equation (10) is *value* (or similarly *policy*) *iteration* [14], which initializes the optimum value of each state randomly or using a guess and updates the value iteratively using the immediate contribution and the expected value of the future states according to the value function until convergence. However, this type of methods only applies to problems whose states are discrete and enumerable. The state space in our model is not discrete, so the iteration over the whole space is not possible.

An approach that avoids the iteration over the whole state space is to use the *value function approximation* [11, 16], in which the value function $V(S)$ is estimated explicitly by representing states with features and learning a model, e.g., linear combinations of features and neural networks, using, e.g., the gradient descent method. While feature extraction from our states can be done, its process is not trivial if we would like use features from the articles involved in a state. In addition, to obtain the target value (e.g., V(S)) for the gradient descent training, an immediate reward is needed as part of the target value. However, in our problem, the reward function is very sparse. Moreover, our environment changes over time where new articles arrive frequently which may cause the state transition function and thus the value function change over time. Re-learning or updating the learned value function online may not be feasible.

Alternatively, it is possible to design a policy function directly by solving an approximation of the problem over a horizon. This class of techniques is called *lookahead policy* in approximate dynamic programming, and is distinguished from the value function approximation in that it uses random samples to simulate the future and make decisions directly based on the simulated future online without explicitly learning a value function.

4.3 Lookahead Paywall Policy

In this section, we propose a solution that uses the *lookahead* technique to estimate the decision function directly. In our solution, we replace the expectation in Equation (10) with an estimation. In particular, we make the following approximations when formulating the model: 1) we uses a short horizon H to limit the number of future time steps to look into, and 2) instead of using the full set of possible outcomes, we use Monte Carlo sampling to select a subset of outcomes starting at time t. Moreover, we use the two-stage approximation for decision making. That is, we assume that we are in a known state S_t at time t; the second stage starts at time $t+1$, where we have different sample paths (i.e., realizations) of the future states from time $t+1$ to $t+H$. Let ω be a sample path of possible article requests from time $t+1$ to $t+H$ (which are stochastic), and $\widetilde{S}_{t'}(\omega)$ and $\widetilde{x}_{t'}(\omega)$ be the state and decision variables at time t' for the sample path ω accordingly (when we are in time t). Decisions are based on all stochastic variables over the horizon t to $t+H$ as follows:

$$X^{\pi^*}(S_t) = \arg \max_{\substack{x_t, \widetilde{x}_{t'}(\omega) \in \{0,1\}, \\ t+1 \le t' \le t+H, \forall \omega \in \Omega}} \{x_t C(S_t, x_t) + (1-x_t) \sum_{\omega \in \Omega} p(\omega) \sum_{t'=t+1}^{t+H} \gamma^{t'-t} \, C(\widetilde{S}_{t'}(\omega), \widetilde{x}_{t'}(\omega))\} \quad (12)$$

where, Ω is the set of sample paths. In fact, at time t, we solve the problem optimally over horizon t to $t+H$ (using sampling) and find $x_t, \widetilde{x}_{t+1}(\omega)$ to $\widetilde{x}_{t+H}(\omega)$. However, we are not interested in values of $\widetilde{x}_{t+1}(\omega)$ to $\widetilde{x}_{t+H}(\omega)$. We are only interested in x_t, which is a decision at time t. After decision x_t is made, we advance through time and the process is repeated. Note that in Equation (12), x_t is common among all realizations. This procedure results in a simple and efficient method which can be applied to each article request in a session online.

Algorithm 1 shows the designed approach for the paywall problem based on Equation (12). This algorithm receives the current state S_t and navigational graph $\mathcal{G}$ and returns either 1 (i.e., show the paywall after this state) or 0 (continue without showing the paywall). Line 6-13 in the algorithm create a sample path (i.e., ω) based on the potential sessions encoded in $\mathcal{G}$. The only modification is that we do not proceed with sampling if the entropy of the current vertex is greater than 0.5. The entropy is calculated based on the probability of adjacent/next articles (determined based on weights of edges) in $\mathcal{G}$ as follows:

$$entropy(a_i) = -\sum_{a_j \in \mathcal{N}_e(a_i, \mathcal{G})} \frac{w_{ij}}{\sum_k w_{ik}} \log_b \frac{w_{ij}}{\sum_k w_{ik}} \quad (13)$$

where w_{ij} is the weight of edge between node a_i and a_j in the navigational graph, and $\mathcal{N}_e(a_i, \mathcal{G})$ is the set of neighbor vertices of a_i in $\mathcal{G}$. Note that we utilize the relative entropy (i.e., the base of log is b, where b is the number of adjacent vertices), so it is always between 0 and 1. The reason for using entropy is that if the article entropy is high, jumping to the next article is likely to introduce some noise (i.e., irrelevant articles). In particular, line 8 samples a next potential article $\hat{a}_{t'}$ from the set of neighbor vertices in $\mathcal{G}$ (i.e., $\mathcal{N}_e(a_t, \mathcal{G})$) based on the distribution $Pr(\mathcal{G}, \mathcal{N}_e(a_t, \mathcal{G}))$. In this distribution, for each vertex the probability of each adjacent vertex a_j is calculated by dividing the weight of the outgoing edge to a_j by the sum of all weights of outgoing edges of the vertex in navigational graph $\mathcal{G}$. Line 16-20 determines the best paywall time for different sample paths of articles by going over the states (line 17) in the path and finding the best contribution (line 18 and 19). Note that we need to calculate the second part of Equation (12) if we assume that x_t is 0, and in case that x_t equals 1, the first term

Algorithm 1: Lookahead Paywall Policy Algorithm

Input: S_t, $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$
Output: x_t
1 $P \leftarrow 0$
2 **for** $i = 1$ *to* $|\Omega|$ **do**
3 $\quad a_t \leftarrow$ Requested article of S_t
4 $\quad \omega \leftarrow []$
5 $\quad t' \leftarrow t+1$, $Stop \leftarrow False$
6 $\quad$ **while** $t' \le t+H$ ***and*** $Stop = False$ **do**
7 $\quad\quad$ **if** $entropy(a_t) \le 0.5$ **then**
8 $\quad\quad\quad \hat{a}_{t'} \sim Pr(\mathcal{N}_e(a_t, \mathcal{G}))$ ▷ $\mathcal{N}_e$ is the set of adjacent vertices for a_t
9 $\quad\quad\quad \omega \leftarrow [\omega, \hat{a}_{t'}]$
10 $\quad\quad\quad a_t \leftarrow \hat{a}_{t'}$
11 $\quad\quad$ **else**
12 $\quad\quad\quad Stop \leftarrow True$
13 $\quad\quad t' \leftarrow t'+1$
14 $\quad P_i \leftarrow 0$
15 $\quad \widetilde{S}_t \leftarrow S_t \; t' \leftarrow t+1$
16 $\quad$ **while** $t' \le t+|\omega|$ **do**
17 $\quad\quad \widetilde{S}_{t'} \leftarrow S^M(\widetilde{S}_{t'-1}, 0, \omega_{t'})$ ▷ $\omega_{t'}$ is the article at time t' in ω
18 $\quad\quad$ **if** $P_i < \gamma^{t'-t} \, C(\widetilde{S}_{t'}, 1)$ **then**
19 $\quad\quad\quad P_i \leftarrow \gamma^{t'-t} \, C(\widetilde{S}_{t'}, 1)$
20 $\quad\quad t' \leftarrow t'+1$
21 $\quad P \leftarrow P + P_i$
22 **if** $C(S_t, 1) > P/|\Omega|$ **then**
23 $\quad$ **return** 1
24 **else**
25 $\quad$ **return** 0

in the equation (12) results in the whole contribution. Given that our sampling is unbiased, we have $p(\omega) = 1/|\Omega|$. Finally, we can determine the best value for x_t by comparing the contributions of two cases ($x_t = 0$ or 1) accordingly (line 22).

Algorithm 2: Adaptive Paywall Algorithm

Input: User Requests, Navigation Graph $\mathcal{G}$
Output: Paywall decision
forall *Requested article* $\hat{a}$ **do**
$\quad$ **if** $t = 0$ **then**
$\quad\quad$ Initialize the state S_0
$\quad\quad$ Show $\hat{a}$
$\quad$ **else**
$\quad\quad S_t = S^M(S_{t-1}, x_{t-1}, \hat{a})$
$\quad\quad x_t \leftarrow$ LookaheadPaywallPolicy(S_t, $\mathcal{G}$)
$\quad\quad$ **if** $x_t = 1$ **then**
$\quad\quad\quad$ Show paywall and terminate the session
$\quad\quad$ **else**
$\quad\quad\quad$ Show $\hat{a}$
$\quad t \leftarrow t+1$

Algorithm 2 illustrates the overall procedure for the adaptive paywall approach. It receives a user request and makes a decision accordingly. Line 3 initializes the session by setting $\overline{u}_0$, $\overline{c}_0$, and a_0 of the first state (i.e., S_0) to $\phi(a_0)$, $\psi(a_0)$ and requested article $\hat{a}$ accordingly. The algorithm always shows the first article (i.e., $x_0 = 0$). If the requested article $\hat{a}$ is not the first one, we first change the state to the next state (line 6), and then based on the result of the policy (Algorithm 1) we either show the article, or go to the paywall and terminate the session.

The above algorithms makes decisions online by simulating future states based on the navigation graph. New articles and changes in user navigation patterns can be incorporated easily by updating the graph. As long as the graph is updated, the algorithms can capture the new changes in the environment.

5 EMPIRICAL STUDY

We applied and evaluated the proposed method on a real dataset from The Globe and Mail[1], a major newspaper in Canada.

5.1 Dataset

The Globe and Mail dataset is collected based on the *omniture*[2] data collection platform. Whenever a user reads an article, watches a video or generally interacts with the news portal, a record called *hit* is generated and stored. Typically, a hit contains information such as user environmental variables (i.e., *browser type*, *IP address*), *user id* (for a subscribed user), *session id*[3], *visited article id*, *click timestamp* as well as *special events* (e.g., subscription, sign in, etc). The original dataset contains about 2 billion hits. We aggregate and roll up the dataset from page view hits to session level (i.e., visit). Moreover, we calculate dwell time (i.e., the time that a user spends on an article) using the difference between two consecutive hits' (e.g., articles) timestamps The last article in the session is ignored since we cannot calculate its dwell time. We use the data collected within 2014-01 to 2014-07 for the evaluation purpose. After preprocessing (i.e., aggregation), the dataset contains 4,913,423 sessions from subscribed users. We use sessions with minimum 10 articles as the test and the rest as the training set (e.g., to build the navigational graph). The test data set provide real sequences of article requests for use to evaluate the proposed method. A minimum of length 10 gives us more possible time points for making paywall decisions in a session. The numbers of sessions in the training and test sets are 4,806,204 and 107,219, respectively.

5.2 Utility and Cost Models

While the utility and cost of an article can be defined differently in our model, we use two intuitive utility and cost models in our experiments, namely, *Engagement to Cost (E/C)* and *Acquisition to Cost (A/C)*. For both models the cost (i.e., ψ) is defined as the article length (in terms of the number of 1KB pages) as the lengthy articles often need more efforts and resources to produce. The utilities for these models are defined as follows:

- The utility of article a in (E/C) model is defined as:

$$\phi(a) = \frac{Total\ dwell\ time\ by\ all\ users\ on\ a}{Total\ number\ of\ visits\ of\ a} \tag{14}$$

where the unit of time is second.

- The utility of article a in (A/C) model is defined as:

$$\phi(a) = \frac{Total\ number\ of\ visits\ of\ a\ before\ subscription}{Number\ of\ visits\ of\ a} \tag{15}$$

where the numerator is the number of subscribers who visited the article before the subscription (i.e., in the subscription session).

Note that the second measure has very small values as the number of subscribers is much smaller than non-subscribed users.

5.3 Baselines and Performance Measures

We compare the proposed Lookahead Paywall Policy (denoted as **LAP**) model with the following baselines:

- **Fixed Policy (FP):** This is a commonly-used paywall mechanism which allows a user to visit a maximum number, T, of articles during a session.
- **Average Threshold (AT):** This is a type of myopic policy [11] which defines the analytical decision function only based on the *current state* using a threshold. The decision function in this method is defined as follows:

$$x_t \triangleq X^{\pi}(S_t|\tau) \triangleq \begin{cases} 1 & \text{If } (\overline{u}_t/\overline{c}_t \leq \tau) \\ 0 & \text{Otherwise} \end{cases} \tag{16}$$

where τ is set as the average ratio of the utility to cost, calculated based on the sessions in the training set.

- **Policy Function Approximation (PFA):** This policy is based on Equation (16), but the parameter τ is optimized using sessions in the training set. Given (16), the policy search in the base optimization problem (8) is changed to a parameter search. Therefore, we can rewrite the optimization function in (8) as:

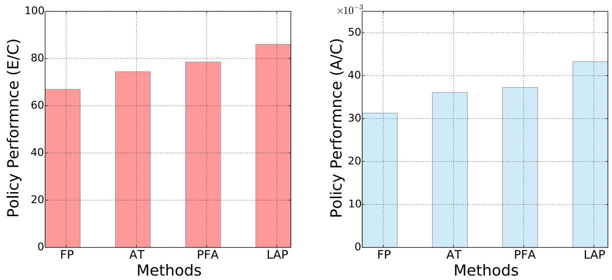
$$\max_{\tau} J(\tau) = \max_{\tau} \mathbf{E}\{\sum_{t=0}^{\infty} \gamma^t\ C(S_t, X^{\pi}(S_t|\tau))\} \tag{17}$$

Finding the best value of τ is a stochastic search problem. However, we cannot compute $J(\tau)$ in a compact form and $X^{\pi}(S_t|\tau)$ is not differentiable. Therefore, we use finite difference gradient [6, 15] which is a common approach for solving Equation (17).

In our experiments, we vary the value for T in the FP method. For our proposed LAP method. We set γ to 1 (which means that the future articles visits have the same value as the current one) and the horizon size H in Algorithm 1 to 4 unless indicated otherwise and sample size to 10 for all experiments.

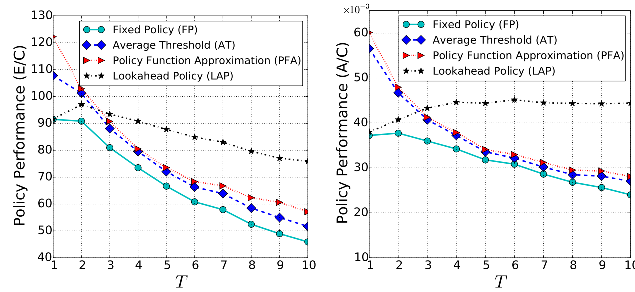
We use the following *performance measures* when comparing the methods. (1) *Policy performance*, which is defined as the average ratio of aggregated utility of all articles in a session to the aggregated cost of all articles in the session over all the tested sessions. The utility of an article is defined using either the E/C or A/C model (according to Equation (14) or (15)). Thus, we have two policy performance measures: E/C based or A/C based. (2) *Policy performance at different percentages of articles delivered to users*, which measures

[1] www.theglobeandmail.com
[2] https://my.omniture.com
[3] In Omniture data collection platform, no activity in 30 minutes is considered as the end of session.



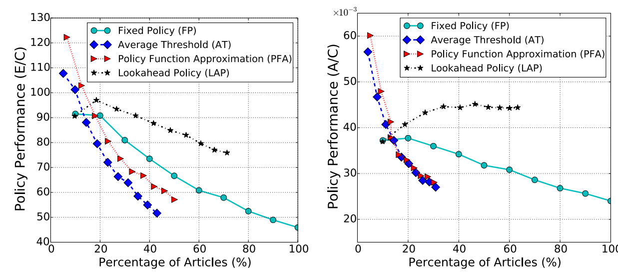
(a) E/C model (b) A/C model

Figure 4: Average policy performance of different methods.



(a) E/C model (b) A/C model

Figure 5: Policy performance for different models.



(a) E/C model (b) A/C model

Figure 6: Policy performance vs. article delivery percentage.

the average utility-to-cost ratio of the sessions when a percentage of articles is presented to users. Obviously, the higher the ratio, the better the performance. (3) *The percentage of active sessions at each time point*, which is the percentage of sessions that have not received the paywall at each time point. A method A with more active sessions at time t than method B is better at t if it has at least the same or better policy performance than B at t. This is because as long as the trade-off between cost and utility is fine, keeping the user active without presenting the paywall can further engage the user and also deliver advertisements.

5.4 Policy Performance Analysis

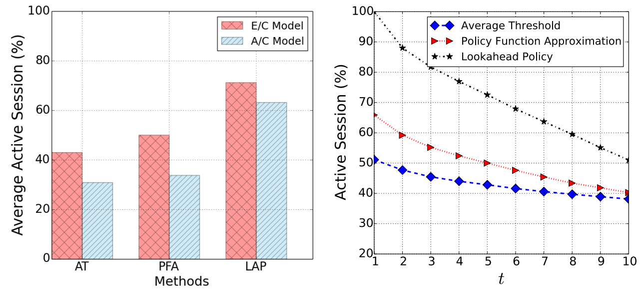
Figure 4 shows the overall average policy performance of the proposed model compared to the baselines for the engagement (i.e., E/C) and acquisition (i.e., A/C) utility models. For the FP method, the result is the average over T values ranging from 1 to 10. The Lookahead Policy model (LAP) shows 28.4 % and 38.3 % performance improvements over the traditional Fixed Policy (FP) for the E/C and A/C model respectively. It also outperforms the Average Threshold (AP) and Policy Function Approximation (PFA) on both E/C and A/C models.

In Figure 5, we compare the policies at different T values, where all policies are allowed to show maximum T articles. Figure 5a illustrates the performance of the polices for the E/C model, which shows that the commonly-used FP policy has the lowest performance for all T values. It also shows that the performances of PFA and AT are better than the other methods at the beginning when T is small (i.e., less than 3). This is because they greedily terminate sessions if the articles requested so far do not look promising based on the threshold τ without looking at the future. On the other hand, if the LAP method is forced to stop early, it does not have the opportunity to make a better decision later on even if it finds one by looking ahead. As T increases, LAP performs better as it can present the paywall at a later time if it thinks it is better than current ones. Another observation is while the performances of all policies for the E/C model decline as we increase T (due to fact that users may visit more engaging articles at the start of sessions), the LAP policy outperforms the other policies by keeping alive the sessions with the promising future. Finally, between PFA and AT, PFA is marginally better than the AT method. Similar observations are found in Figure 5b, which illustrates the performances of the policies for the A/C model at different T values. The only difference is that as T increases, LAP's performance does not decline, which means that the LAP policy successfully navigates users to the articles that are useful for user acquisition (i.e., those which have been visited by the converted users and have good utility-to-cost ratios).

5.5 Performance vs. Delivery Percentage

In this section, we study the performance of each policy at different percentages of delivered articles (similar to precision at each recall level in information retrieval). To do this, for each policy and each test session s and at each time point t (from 1 to 10), we compute the percentage of delivered articles by time t as the number of articles delivered by the policy by time t divided by the total number of articles in session s, and also compute the ratio of the aggregated utility to the aggregated cost of all the delivered articles in the session by time t. In this way, for each policy and each session we obtain 10 ⟨delivered article percentage, policy performance⟩ pairs, one for each time point. We finally take an average of each pair over all the test sessions for the policy. Thus, each policy has 10 averaged pairs, one for each time point. For the FP method, we vary T from 1 to 10 to collect the data for each time point.

Figure 6a shows the policy performance against the percentage of delivered articles for the E/C model. As can be observed, at the small article percentage levels (less than 15%), PFA and AT are better than LAP and FP. However, as we increase the percentage of articles, both LAP and FP outperform PFA and AT, with LAP being



(a) Average active session **(b) E/C model** **(c) A/C model**

Figure 7: Active session for different utility to cost models.

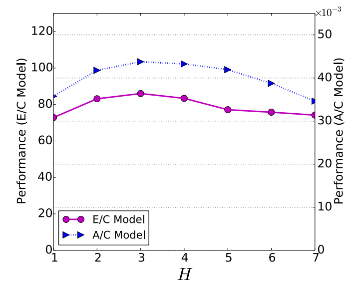


Figure 8: performance for different H.

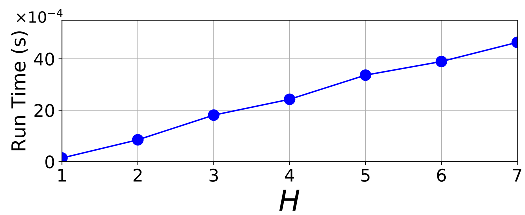


Figure 9: Average Runtime per request.

significantly better than all other polices consistently. In particular, LAP performs 31.0 % better than FP when 70 % of articles are shown to users. Similar results are found in Figure 6b, which shows the policy performance against the percentage of delivered articles for the A/C model. For example, the LAP performance is 48.0 % better than that of FP when 63 % of articles are shown to the users.

5.6 The Effect of Policies on Users' Sessions

In this section, we investigate the effect of different policies on the number of active sessions (i.e., the percentage of sessions/users which have not received paywall before time point t). Intuitively, an online news agency prefers to have a high number of active readers all the time as long as the articles the reader reads have good utilities and their total cost is reasonable (i.e., the ratio of utility to cost is acceptable). Keeping a user active without presenting the paywall can further engage the user and allow more display of advertisements. Figure 7a illustrates the average percentage of active sessions (averaged over different time points in a session) for both E/C and A/C models. As can be seen, LAP has more active sessions on average compared to other policies. For example, the average number of active sessions of LAP is 42.5 % and 66.0 % more than PFA and AT methods respectively for the E/C model. Note that we did not show the FP method in this study because all sessions are active till time T in FP (which is assumed to be 10). Figure 7b and 7c illustrate the percentage of active sessions at different t values. PFA and AT terminate many sessions at the beginning by showing the paywall because close to half of the articles do not meet their threshold which is the average or close to average utility-to-cost ratio learned from the training data, while in LAP there are more active sessions at all the time points and the percentage of active sessions decreases more smoothly. Considering that LAP has better utility-to-cost ratios than other methods at almost all the time points (as shown in Figure 5), LAP is better as it has more active users without sacrificing the utility-to-cost ratio.

5.7 Sensitivity and Runtime Analysis

We analyze the performance sensitivity of LAP policy with different horizon sizes (i.e., H). We change the horizon size H and calculate the policy performance accordingly. Figure 8 shows the effect of horizon size H on the LAP performance for the E/C and A/C models. As can be observed, by incrementing H, the performance increases slightly for both E/C and A/C models, becomes stable in range of 2 to 4, and then starts to decline. This suggests that any value in range 2 to 4 would be a good choice for this parameter. Figure 9 shows the average run time per request for the LAP method. The experiments run on 2.2 GHz Intel Core i7 machine with macOS operating system. As can be seen, the response time per request is fast (a few milliseconds) and increases almost linearly with H.

6 CONCLUSION AND FUTURE WORK

We proposed an adaptive paywall mechanism for digital news media. The traditional paywall model allows a user to see a fixed number of articles and then directs them to the subscription page. We argued that this approach does not lead to more subscriptions and sacrifices other business objectives (e.g., increasing the user dwell time, the number of visits, etc.). We proposed a solution by formulating the paywall problem as a sequential decision problem that optimizes the ratio of the aggregated utility of the articles presented to the user to their aggregated cost. We defined our problem and its components in an approximate dynamic programming paradigm, analyzed possible ways to solve it, and proposed a solution that uses a data-driven lookahead policy. We applied the proposed method to a real dataset from a national newspaper in Canada and showed the benefits and superiority of the method over the existing method and other baselines. To the best of our knowledge, the adaptive paywall problem has not been studied previously.

There are interesting directions for future work. The current evaluation was done in an off-line setting to provide a proof of concept. Before the system can be deployed, an online evaluation in a real-world environment will be conducted. Another challenge is how to define the utility of a new article. In our experiment, we used the dwell time or the number of visits to the article in the browsing history. For new articles, such information does not exist. We are working on how to predict the utility of an article based on its content and other information. We will also investigate how to model the dynamics of utility and cost over time and incorporate it into the proposed model. Finally, we hope what we have presented in this paper can inspire others searching for solutions to similar problems.

ACKNOWLEDGMENTS

This work is funded by Natural Sciences and Engineering Research Council of Canada (NSERC), The Globe and Mail, and the Big Data Research, Analytics and Information Network (BRAIN) Alliance established by the Ontario Research Fund - Research Excellence Program (ORF-RE). We would also like to thank Dr. Ricky Fok and Dr. Manos Papagelis for the helpful discussions at the early stage of this work.

REFERENCES

[1] Mark Broadie, Paul Glasserman, and Zachary Ha. 2000. Pricing American options by simulation using a stochastic mesh with optimized weights. In *Probabilistic Constrained Optimization*. Springer, 26–44.

[2] Julia Cagé. 2016. *Saving the media : capitalism, crowdfunding, and democracy.* The Belknap Press of Harvard University Press, Cambridge, Massachusetts.

[3] Han Cai, Kan Ren, Weinan Zhang, Kleanthis Malialis, Jun Wang, Yong Yu, and Defeng Guo. 2017. Real-Time Bidding by Reinforcement Learning in Display Advertising. In *Proceedings of WSDM'17*. ACM, 661–670.

[4] Hsiang Chyi. 2005. Willingness to pay for online news: An empirical study on the viability of the subscription model. *Journal of Media Economics* 18, 2 (2005), 131–142.

[5] Heidar Davoudi, Morteza Zihayat, and Aijun An. 2017. Time-Aware Subscription Prediction Model for User Acquisition in Digital News Media. In *Proceedings of the 2017 SIAM International Conference on Data Mining*. SIAM, 135–143.

[6] Peter Deisenroth, Gerhard Neumann, Jan Peters, et al. 2013. A survey on policy search for robotics. *Foundations and Trends in Robotics* 2, 1–2 (2013), 1–142.

[7] Richard Fletcher and Rasmus Kleis Nielsen. 2017. Paying for Online News: A comparative analysis of six countries. *Digital Journalism* 5, 9 (2017), 1173–1191.

[8] Manuel Goyanes. 2014. An empirical study of factors that influence the willingness to pay for online news. *Journalism Practice* 8, 6 (2014), 742–757.

[9] Merja Myllylahti. 2014. NEWSPAPER PAYWALLS-THE HYPE AND THE REALITY: A study of how paid news content impacts on media corporation revenues. *Digital journalism* 2, 2 (2014), 179–194.

[10] Nic Newman, Richard Fletcher, Antonis Kalogeropoulos, David AL Levy, and Rasmus Kleis Nielsen. 2017. Reuters Institute Digital News Report 2017. (2017).

[11] Warren B Powell. 2011. *Approximate Dynamic Programming: Solving the curses of dimensionality.* John Wiley & Sons.

[12] Guy Shani, David Heckerman, and Ronen I Brafman. 2005. An MDP-based recommender system. *Journal of Machine Learning Research* 6 (2005), 1265–1295.

[13] Richard Sutton. 1988. Learning to predict by the methods of temporal differences. *Machine learning* 3, 1 (1988), 9–44.

[14] Richard Sutton and Andrew Barto. 2011. Reinforcement learning: An introduction. (2011).

[15] Richard S Sutton, David A McAllester, Satinder P Singh, and Yishay Mansour. 2000. Policy gradient methods for reinforcement learning with function approximation. In *Advances in neural information processing systems*. 1057–1063.

[16] Gavin Taylor and Ronald Parr. 2009. Kernelized value function approximation for reinforcement learning. In *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM, 1017–1024.

[17] Christopher JCH Watkins and Peter Dayan. 1992. Q-learning. *Machine learning* 8, 3-4 (1992), 279–292.

[18] A. Williams. 2016. Paying for Digital News: The rapid adoption and current landscape of digital subscriptions at US newspapers. *American Press Inst.* (2016).